```
Last login: Sun Feb 13 13:21:49 2022 from broadband-46-242-8-209.ip.moscow.rt.ru
[student897_11@bigdataanalytics-worker-3 ~]$ cqlsh
Connected to Test Cluster at 127.0.0.1:9042
[cqlsh 6.0.0 | Cassandra 4.0.1 | CQL spec 3.4.5 | Native protocol v5]
Use HELP for help.
cqlsh> create keyspace keyspace_sapr with replication  = {'class': 'SimpleStrategy', 'replication_factor':1};
cqlsh> USE
cqlsh> USE keyspace_sapr
   ...
   ...
cqlsh> USE keyspace_sapr;
cqlsh:keyspace_sapr> create table sapr_shtatka
                 ... (id int,
                 ... name text,
                 ... dolzhnost text,
                 ... salary float
                 ... primary key (id));
SyntaxException: line 6:12 mismatched input '(' expecting ')' (...,salary floatprimary key [(]...)
cqlsh:keyspace_sapr> create table sapr_shtatka
                 ... (id int,
                 ... name text,
                 ... dolzhnost text,
                 ... salary float,
                 ... primary key (id));
cqlsh:keyspace_sapr> select keyspace_name, table_name from system_schema.tables where keyspace_name = 'keyspace_sapr';

 keyspace_name | table_name
---------------+--------------
 keyspace_sapr | sapr_shtatka

(1 rows)
cqlsh:keyspace_sapr>
```

```
(1 rows)
cqlsh:keyspace_sapr> insert into sapr_shtatka(id, name, dolzhnost,salary)
                 ... values(1,'Nikita','analyst',100000);
cqlsh:keyspace_sapr> insert into sapr_shtatka(id, name, dolzhnost,salary)
                 ... values(2,'Aleksey','system analyst',150000);
cqlsh:keyspace_sapr> insert into sapr_shtatka(id, name, dolzhnost,salary)
                 ... values(3,'Ivan','business analyst',140000);
cqlsh:keyspace_sapr>
cqlsh:keyspace_sapr> insert into sapr_shtatka(id, name, dolzhnost,salary)
                 ... values(4,'Natali','buhgulter',140000);
cqlsh:keyspace_sapr>
cqlsh:keyspace_sapr> insert into sapr_shtatka(id, name, dolzhnost,salary)
                 ... values(5,'Natali','ingener',80000);
cqlsh:keyspace_sapr> insert into sapr_shtatka(id, name, dolzhnost,salary)
                 ... values(6,'Petr','economist',90000);
cqlsh:keyspace_sapr> select * from sapr_shtatka;

 id | dolzhnost        | name    | salary
----+------------------+---------+---------
  5 |          ingener |  Natali |   80000
  1 |          analyst |  Nikita |   1e+05
  2 |   system analyst | Aleksey | 1.5e+05
  4 |        buhgulter |  Natali | 1.4e+05
  6 |        economist |    Petr |   90000
  3 | business analyst |    Ivan | 1.4e+05

(6 rows)
```

```
>>> shtatka_df = spark.read \
...     .format("org.apache.spark.sql.cassandra") \
...     .options(table="sapr_shtatka", keyspace="keyspace_sapr") \
...     .load()
>>> shtatka_df.printSchema()
root
 |-- id: integer (nullable = true)
 |-- dolzhnost: string (nullable = true)
 |-- name: string (nullable = true)
 |-- salary: float (nullable = true)

>>>
```

```
>>>
>>> shtatka_df.show()
+---+---------------+-------+--------+
| id|       dolzhnost|   name|  salary|
+---+---------------+-------+--------+
|  5|         ingener| Natali| 80000.0|
|  1|         analyst| Nikita|100000.0|
|  2|  system analyst|Aleksey|150000.0|
|  6|       economist|   Petr| 90000.0|
|  4|       buhgulter| Natali|140000.0|
|  3|business analyst|   Ivan|140000.0|
+---+---------------+-------+--------+

>>>
```

```
>>>
>>> shr_df = spark.sql("""select 7 as id, "Dima" as name, "osinizator" as dolzhnost, 50000 as salary""")
>>> shr_df.printSchema()
root
 |-- id: integer (nullable = false)
 |-- name: string (nullable = false)
 |-- dolzhnost: string (nullable = false)
 |-- salary: integer (nullable = false)

>>> shr_df.show()
+---+----+----------+------+
| id|name| dolzhnost|salary|
+---+----+----------+------+
|  7|Dima|osinizator| 50000|
+---+----+----------+------+

>>>
```

```
>>> shtatka_df.show()
+---+---------------+-------+--------+
| id|       dolzhnost|   name|  salary|
+---+---------------+-------+--------+
|  5|         ingener| Natali| 80000.0|
|  1|         analyst| Nikita|100000.0|
|  2|  system analyst|Aleksey|150000.0|
|  6|        economist|   Petr| 90000.0|
|  4|       buhgulter| Natali|140000.0|
|  7|      osinizator|   Dima| 50000.0|
|  3|business analyst|   Ivan|140000.0|
+---+---------------+-------+--------+

>>>
```

```
>>> shtatka_df.filter(F.col("id")=="7").show()
+---+----------+----+-------+
| id| dolzhnost|name| salary|
+---+----------+----+-------+
|  7|osinizator|Dima|50000.0|
+---+----------+----+-------+

>>>
```

```
>>> shtatka_df.groupby("salary").sum().show()
+--------+-------+-----------+
|  salary|sum(id)|sum(salary)|
+--------+-------+-----------+
|100000.0|      1|   100000.0|
| 90000.0|      6|    90000.0|
| 80000.0|      5|    80000.0|
|140000.0|      7|   280000.0|
| 50000.0|      7|    50000.0|
|150000.0|      2|   150000.0|
+--------+-------+-----------+
```

```
>>>
>>> shtatka_df.groupby(["dolzhnost","salary"]).sum().show()
+----------------+--------+-------+-----------+
|       dolzhnost|  salary|sum(id)|sum(salary)|
+----------------+--------+-------+-----------+
|  system analyst|150000.0|      2|   150000.0|
|        economist| 90000.0|      6|    90000.0|
|      osinizator| 50000.0|      7|    50000.0|
|business analyst|140000.0|      3|   140000.0|
|         analyst|100000.0|      1|   100000.0|
|         analyst|120000.0|      8|   120000.0|
|         ingener| 80000.0|      5|    80000.0|
|       buhgulter|140000.0|      4|   140000.0|
+----------------+--------+-------+-----------+

>>>  shtatka_df.groupby(["dolzhnost",]).sum("salary").show()
  File "<stdin>", line 1
    shtatka_df.groupby(["dolzhnost",]).sum("salary").show()
    ^
IndentationError: unexpected indent
>>> shtatka_df.groupby(["dolzhnost"]).sum("salary").show()
+----------------+-----------+
|       dolzhnost|sum(salary)|
+----------------+-----------+
|        economist|    90000.0|
|         ingener|    80000.0|
|  system analyst|   150000.0|
|business analyst|   140000.0|
|         analyst|   220000.0|
|       buhgulter|   140000.0|
|      osinizator|    50000.0|
+----------------+-----------+

>>>
```

```
IndentationError: unexpected indent
>>> shtatka_df.filter(F.col("salary")==100000).explain()
== Physical Plan ==
*(1) Filter (isnotnull(salary#33) && (salary#33 = 100000.0))
+- *(1) Scan org.apache.spark.sql.cassandra.CassandraSourceRelation@33aa8011 [id#30,dolzhnost#31,name#32,salary#33] PushedFilters: [IsNotNull(salary), EqualTo(salary,100000.0)], ReadSchema: struct<id:int,dolzhnost
:string,name:string,salary:float>
>>>
```